



OPEN

N^1 -methyl-pseudouridine is incorporated with higher fidelity than pseudouridine in synthetic RNAs

Tien-Hao Chen, Vladimir Potapov, Nan Dai, Jennifer L. Ong & Bijoyita Roy✉

In vitro transcribed synthetic messenger RNAs (mRNAs) represent a novel therapeutic modality. To overcome the inherent immunogenicity, as well as to increase the therapeutic efficacy of the molecules, uridine analogs—such as pseudouridine (Ψ) and N^1 -methyl-pseudouridine (m1 Ψ), are incorporated in the synthetic mRNA. To decipher the fidelity with which these modifications are incorporated during the *in vitro* transcription (IVT) process, we compared the incorporation fidelity of uridine analogs with different RNA polymerases. We demonstrate that m1 Ψ is incorporated with higher fidelity than Ψ . The fidelity of nucleotide incorporation differs between RNA polymerases; however, the spectrum of mutations observed between the RNAPs is similar. We also show that the array of nucleotide misincorporation is not dependent on the template DNA sequence context and that the distribution of these misincorporated nucleotides is not localized to any specific region along the length of the RNA. Based on our findings, we introduce a novel method to improve uridine analog incorporation fidelity during IVT. Our proof-of-concept experiments for higher-fidelity incorporation of uridine analogs during IVT provide guidelines when choosing RNAPs for the generation of modified uridine-containing mRNAs *in vitro*.

Synthetic messenger RNAs (mRNAs) are a novel modality for vaccines, and they are also being evaluated as a vector for therapeutics^{1–3}. Despite there being several advantages over conventional protein-based approaches, mRNA-based therapeutics are still in early stages of development. Instability of the synthetic mRNAs and the immune responses generated against these synthetic molecules have been key hurdles in the adaptation of this technology, particularly for therapeutic applications where prolonged expression from the synthetic molecule is desirable and repeated dosing of the drug product is required^{1,2}. The use of chemically modified bases in synthetic mRNAs is an innovation that has allowed for both an ameliorated immune response to the synthetic molecules and increased protein expression from the mRNA, thereby providing an unprecedented opportunity to use synthetic mRNAs as a novel class of therapeutics for a wide range of indications, including two approved vaccines against COVID-19^{4,5}. It has been shown that the incorporation of pseudouridine (ψ), N^1 -methyl-pseudouridine (m1 ψ), 5-methylcytosine (m5C), N6-methyladenosine (m6A) and 2-thiouridine (s2U) into synthetic mRNAs results in reduced immune responses and increased protein expression *in vivo*^{6–14}. Pseudouridine-modified mRNAs have been shown to result in reduced activation of 2'-5'-oligoadenylate synthetase (OAS), RNA-dependent protein kinase (PKR), and toll-like receptors^{6,7,10}. Investigation of ψ derivatives with improved pharmacological properties led to the identification of m1 ψ , the current benchmark for synthetic mRNA-based applications. m1 ψ is a naturally occurring modification found in 18S rRNA and tRNAs^{15–17}, and similar to ψ -modified synthetic mRNAs, the presence of m1 ψ in synthetic mRNAs has been demonstrated to result in reduced activation of RNA sensors in cells^{8,12,14,18}. Furthermore, the presence of m1 ψ in synthetic mRNAs show increased translation efficiency in cell-free extracts, multiple mammalian cell lines, and mouse models^{8,9,12–14,18–20}. The exact mechanism of how m1 ψ enhances translation is not well understood but it has been demonstrated that presence of m1 ψ , alone or in combination with other chemical modifications, can alter ribosome transit time on the modified mRNA, and can increase the mRNA half-life by altering the secondary structure of the synthetic mRNA^{7,14,21,22}.

Current methods to incorporate modified nucleotides in synthetic mRNAs include complete substitution of the standard nucleotide with a chemically modified nucleotide during the process of *in vitro* transcription (IVT) by single-subunit DNA-dependent RNA polymerases (ssRNAPs; such as T7, T3, and SP6 RNAP)¹. In contrast to endogenous mRNAs, in which modified nucleotides occur in specific positions in the mRNA^{23,24}, in synthetic

RNA and Genome Editing, New England Biolabs Inc., Beverly, MA 01915, USA. ✉email: broy@neb.com

mRNAs, the modified nucleotide is present at almost every position where the naturally occurring nucleobase would be. This complete substitution approach is preferred from a regulatory perspective because it results in less molecule-to-molecule variation in the positions of the modified nucleotides along the synthetic mRNA. The exact implications of incorporating the chemical modifications throughout the body of the mRNA is still under investigation. For attaining the maximum effectiveness from the drug product and to ensure that the expression from the synthetic molecule is optimal, it is critical that the modified nucleotide is incorporated in the right place, be compatible with the functional elements of the mRNA, and does not alter the biological function of the synthetic mRNA. Numerous studies have demonstrated that ssRNAPs can incorporate chemically modified nucleotides into RNA^{7,10,25}, but it is unclear whether all ss-RNAPs incorporate the modified nucleotides with comparable fidelity.

We previously established a Pacific Biosciences Single Molecule Real-Time (SMRT) sequencing-based assay to determine the combined transcription and reverse transcription errors and the effects of RNA modifications²⁶. In the previous study, T7 RNAP exhibited higher combined error rates in the synthesis of ψ -, m6A- and 5-hydroxymethylcytidine (hm5C)-modified RNAs, with increased misincorporation of ψ across from dT templated bases. In light of the widespread use of m1 ψ for mRNA vaccines, in the current study, we focused on understanding the fidelity of incorporation of m1 ψ during in vitro transcription using multiple ssRNAPs. The fidelity with which m1 ψ is incorporated during in vitro transcription is a timely question to address because both the SpikeVax (mRNA-1273) and Comirnaty (BNT162b2) SARS-CoV2 mRNA vaccines are synthesized by substituting uridine with m1 ψ throughout the body of the mRNA and both have demonstrated efficacy greater than 94%^{4,5}. We investigated the fidelity of m1 ψ incorporation with three commonly used ssRNAPs (T7, T3 and SP6 RNAPs) and compared it to the fidelity with which uridine and ψ are incorporated. Using four different RNA sequences, including two functional mRNA sequences, we analyzed the effects of sequence context in promoting misincorporation of uridine analogs and identified rA \rightarrow rU substitution errors as the predominant error when uridine analogs (m1 ψ and ψ) are present in the reaction. Finally, based on the nature of the substitution errors observed, we combine sequence optimization of the synthetic mRNA together with an altered RNA synthesis process to reduce the uridine analog incorporation error during in vitro transcription. These proof-of-concept experiments establish a novel method to synthesize mRNAs with improved fidelity of uridine analog incorporation and provide considerations for choosing ssRNAPs for the generation of modified nucleotide-containing mRNAs in vitro.

Results

T7 and SP6 RNAPs incorporate m1 ψ efficiently. The efficiency of m1 ψ incorporation during in vitro transcription was investigated in four different RNA substrates of varying length and sequence. The base composition of the synthesized RNA was analyzed with ultra-high performance liquid chromatography coupled with mass spectrometry. For the long synthetic RNAs with length ranging from 1122 to 4178 nucleotides, the integrity of the RNA was determined using Bioanalyzer. Synthesis of full-length RNAs of expected sizes were observed in reactions performed in the presence of m1 ψ with both T7 RNAP and SP6 RNAP (Supplementary Fig. 1a). Furthermore, similar to ψ , total m1 ψ TP incorporation was as expected in RNA sequences RNA1 (a 1122 nucleotide synthetic sequence that includes all possible four-base combinations)²⁷ and Cypridina luciferase mRNA sequence (CLuc mRNA) when in vitro transcription was performed with T7 RNAP (Supplementary Fig. 1b,c). In order to determine that the RNA synthesized in presence of m1 ψ is indeed of expected length, we subjected two short RNAs, 30-nucleotide RNA and 60-nucleotide long, to intact mass spectrometry analyses (Supplementary Fig. 2). Irrespective of which uridine analog was present in the reaction, the mass of the predominant species observed in reactions performed with T7 RNAP correspond to the run-off transcripts and few non-templated additions were observed²⁸ suggesting that similar to ψ and uridine, m1 ψ is incorporated efficiently during in vitro transcription and the modifications did not disrupt the synthesis of full-length run-off products.

m1 ψ is incorporated with higher fidelity than ψ by T7 RNAP. To determine the fidelity of m1 ψ incorporation during in vitro transcription, we adapted the Pacific Biosciences Single Molecule Real-Time (SMRT) sequencing-based assay (Supplementary Fig. 3) to the PacBio Sequel I system to enable higher sequencing capacity and multiplexing²⁶. We analyzed the errors in the first strand that stem from combined RNAP and RT error, referred hereafter as combined errors. We first determined the combined errors in two synthetic sequences (RNA1 and RNA2) that represent all possible four-base combinations (templates described as DNA-1 and DNA-2 previously²⁶). The error rates of pooled RNA1 and RNA2 (referred as RNA1/RNA2) in reactions with canonical uridine, using the Sequel I system was observed to be $6.4 \pm 0.4 \times 10^{-5}$ error/base as compared to $5.6 \pm 0.8 \times 10^{-5}$ error/base using the PacBio RSII system (Supplementary Table 1)²⁶. Since the combined error rates were comparable between the two platforms, we used the Sequel I system for the subsequent experiments.

We next tested the error rates in reactions with m1 ψ for the two artificial RNA sequences (RNA1 and RNA2), CLuc mRNA as well as BNT162b2 sequences when in vitro transcription reactions are performed with T7 RNAP under standard high-yield reaction conditions. The combined error rates in ψ -modified RNAs were observed to be two-fold higher than unmodified RNAs ($1.3 \pm 0.2 \times 10^{-4}$ errors/base for ψ -containing reactions), while that in m1 ψ -modified RNAs were comparable to unmodified RNAs ($7.4 \pm 0.7 \times 10^{-5}$ error/base for m1 ψ -containing reactions) (Fig. 1a and Table 1).

In order to understand whether the nature of the errors that are introduced when the reactions are performed with different uridine analogs is similar, we compared the error profile observed in the four RNA sequences. Irrespective of the nature of the uridine analog used in the reaction, base substitution was observed to account for the predominant errors ranging from 86 to 95% of total errors for all four RNA sequences (Fig. 1b and Table 1). To further understand if a specific substitution is more prevalent than another and if there are differences when

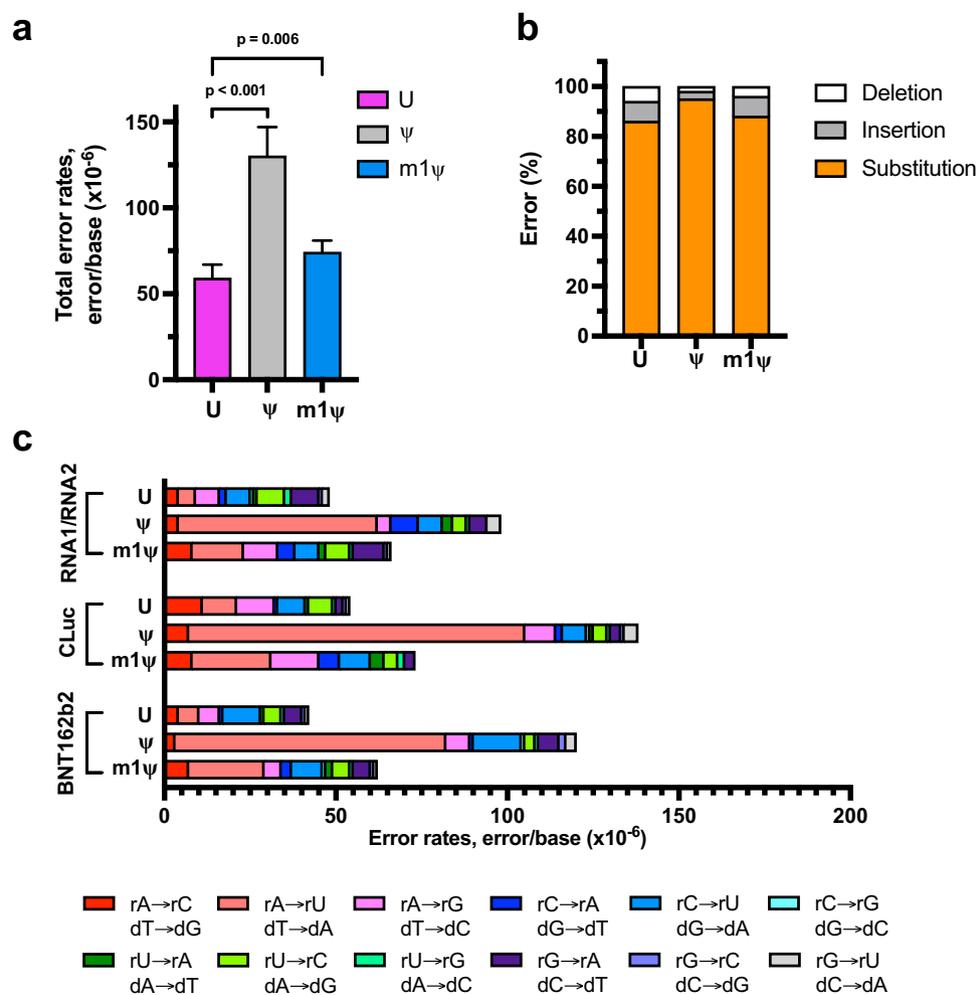


Figure 1. Uridine, ψ , and m1 ψ are incorporated by T7 RNA polymerase with varying error rates during in vitro transcription. The RNA templates synthesized with T7 RNA polymerase were converted into cDNA with ProtoScript II reverse transcriptase followed by library preparation and sequencing. First-strand errors represented here stem from the combination of RNA polymerase and reverse transcriptase errors. (a) Combined first strand error rates are an average of four different RNA sequences (RNA1 and RNA2 are artificial RNA sequences that have been permuted to include every four-base combination; CLuc and BNT162b2 are functional mRNAs), and unpaired t-tests were used to calculate p-values. (b) Distribution of substitution, deletion, and insertion error percentages from four different RNA sequences. (c) Base substitution error profile observed for unmodified and modified RNA sequences are represented. Colors indicate specific substitution errors observed. Polymerase substitution errors are represented as the equivalent T7 RNA polymerase substitution (top substitution) or ProtoScript II reverse transcriptase substitution (bottom substitution).

Base	Total error rate ($\times 10^{-6}$, errors/base)	Substitution (%)	Deletion (%)	Insertion (%)	Total sequenced bases
U	59 \pm 8	86	8	6	45,041,484
ψ	130 \pm 17	95	3	2	14,714,347
m1 ψ	74 \pm 7	88	8	4	9,737,905

Table 1. Total and specific error rates of first strand cDNA in RNA sequences synthesized with T7 RNA polymerase. Total error rates observed in four different RNA sequences (RNA1, RNA2, CLuc and BNT162b2) were pooled. The mean and standard deviation were calculated from six independent measurements (n = 1 for RNA1, n = 1 for RNA2, n = 2 for CLuc mRNA and BNT162b2).

m1 ψ is present in the reaction, we analyzed the substitution profile (Fig. 1c and Supplementary Table 2). For comparing the substitution error profiles, RNA1 and RNA2 data was pooled together to represent substitution errors from artificial RNA sequences and CLuc and BNT162b2 were analyzed separately to understand the substitution error profile of functional mRNAs. For unmodified RNA, no distinct substitution errors were observed for any of the sequences (Fig. 1c and Supplementary Table 2). In contrast, a significant increase in rA → rU/dT → dA

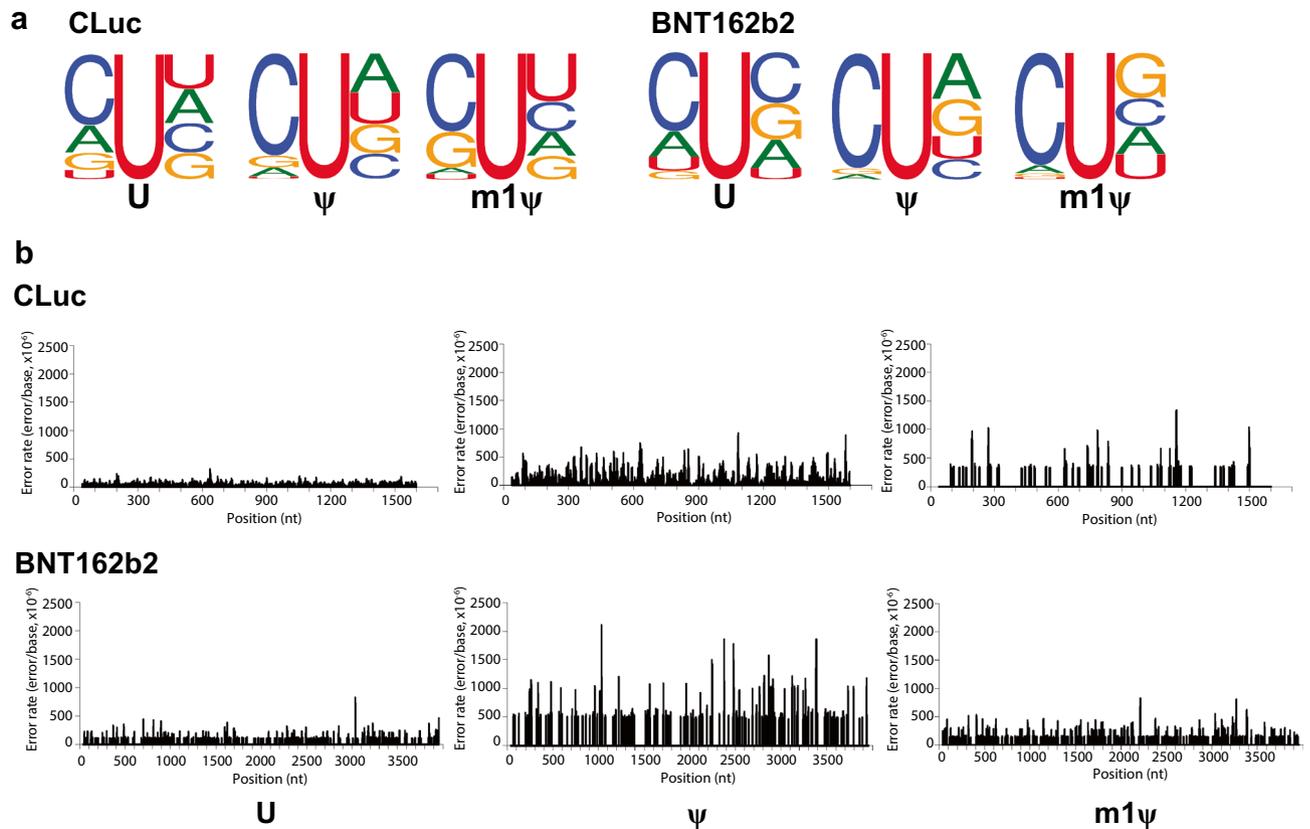


Figure 2. $rA \rightarrow rU/dT \rightarrow dA$ substitution errors observed in ψ -containing RNAs have sequence context preference and occur throughout the transcript body when reactions are performed with T7 RNA polymerase. The sequence context preference and distribution of the $rA \rightarrow rU/dT \rightarrow dA$ substitution errors observed in CLuc and BNT162b2 mRNAs are represented. (a) Logo of sequence surrounding the $rA \rightarrow rU/dT \rightarrow dA$ substitution sites. Seven nucleotides upstream and downstream of the substitution site were examined and for simplicity the immediate sequence upstream and downstream of the substitution site is represented. Sequence logos were built using WebLogo software⁴¹ and the numbers of sites plotted are summarized in Supplementary Table 15. (b) Substitution error distribution along the length of the transcript.

substitutions was observed when reactions were performed with either m1 ψ or ψ (Fig. 1c and Supplementary Table 2), likely due to rm1 ψ TP incorporated in place of rATP (opposite dT) by T7 RNA polymerase as it is less likely that m1 ψ would have a global effect on the incorporation of other bases. The $rA \rightarrow rU/dT \rightarrow dA$ error rates in ψ -modified RNAs were about nine- to 12-fold higher than unmodified RNAs. The $rA \rightarrow rU/dT \rightarrow dA$ substitution was also apparent (up to three-fold greater than unmodified RNAs) in m1 ψ -modified RNAs but was observed to be less compared to ψ -modified counterpart (Fig. 1c and Supplementary Table 2).

Even though the nature of the base substitution errors were similar for all four RNA sequences, we further investigated if there is a specific sequence context within these RNA sequences that have a higher propensity of association with the $rA \rightarrow rU/dT \rightarrow dA$ substitution errors observed. Neighboring sequence analyses of the $rA \rightarrow rU/dT \rightarrow dA$ sites demonstrated a pronounced 5' rC sequence to be present when the $rA \rightarrow rU/dT \rightarrow dA$ substitution was observed in the ψ -modified RNAs (Fig. 2a and Supplementary Fig. 4a). Furthermore, the $rA \rightarrow rU$ sites were observed to be distributed throughout the RNA and specific hot spots for the substitutions were not observed for any of the RNA sequences, suggesting that the errors observed do not have a sequence context dependency (Fig. 2b and Supplementary Fig. 4b).

SP6 RNAP incorporates m1 ψ with higher fidelity than ψ ; overall error rates in reactions performed with SP6 RNAPs are higher than those performed with T7 RNAP.

We next sought to compare different RNAPs to determine if m1 ψ is incorporated with varied fidelity by different ssRNAPs and if the differences observed in error rates in ψ - and m1 ψ -containing RNAs synthesized with T7 RNAP are also observed with other ssRNAPs, T3 and SP6 RNAPs. SP6 RNAP shares 32% identity to T7 RNAP and is also used for generating synthetic mRNAs for therapeutic applications²⁵. On the other hand, T3 RNAP is 82% identical to T7 RNAP. The total combined error rates observed in reactions performed with T3 RNAP were comparable to reactions performed with T7 RNAP suggesting that these two closely related RNAPs have similar fidelity profile (Supplementary Table 3). On the other hand, comparison of the total combined error rates observed from modified and unmodified RNAs synthesized with SP6 RNAP demonstrated one-fold to two-fold higher error rates than those observed with T7 RNAP (Figs. 1a and 3a; Table 1; Supplementary Tables 3 and 4). For unmodified

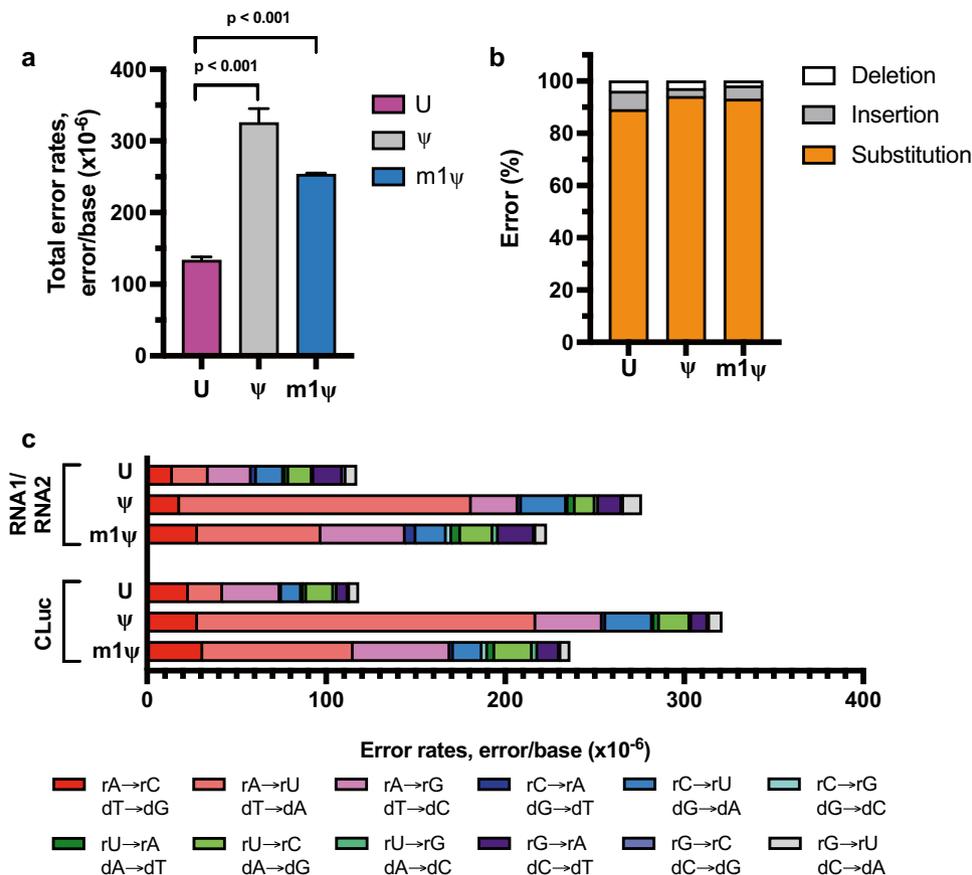


Figure 3. Uridine, ψ , and m1 ψ are incorporated by SP6 RNA polymerase with varying error rates during *in vitro* transcription. The RNA templates synthesized with SP6 RNA polymerase were converted into cDNA with ProtoScript II reverse transcriptase followed by library preparation and sequencing. First-strand errors represented here stem from the combination of RNA polymerase and reverse transcriptase errors. **(a)** Combined first strand error rates are an average of three different RNA sequences (RNA1 and RNA2 are artificial RNA sequences that have been permuted to include every four-base combination; CLuc is a functional mRNA) and unpaired t-tests were used to calculate p-values. **(b)** Distribution of substitution, deletion, and insertion error percentages from three different RNA sequences (RNA1, RNA2, and CLuc). **(c)** Base substitution error profile observed for unmodified and modified RNA sequences are represented. Colors indicate specific substitution errors observed. Polymerase substitution errors are represented as the equivalent SP6 RNA polymerase substitution (top substitution) or ProtoScript II reverse transcriptase substitution (bottom substitution).

RNA (three different sequences), the total combined error rates observed when reactions were performed with SP6 RNAP under standard high-yield reaction conditions were greater than that of T7 RNAP with combined error rates of $1.3 \pm 0 \times 10^{-4}$ errors/base, (Fig. 3a and Supplementary Table 4). For ψ -modified sequences, the error rate was observed to be $3.3 \pm 0.2 \times 10^{-4}$ errors/base. The total combined error rates of m1 ψ -containing RNAs were $2.5 \pm 0 \times 10^{-4}$ errors/base. Similar to T7 RNAP, combined error rates followed the same trend— ψ -modified RNAs demonstrating highest error rates as compared to m1 ψ -modified RNAs and uridine-containing RNAs. (Figs. 1a and 3a; Table 1; Supplementary Table 4).

Similar to T7 RNAPs, base substitution errors were observed to be the most predominant error type ranging between 89 and 94% of total errors when reactions were performed with SP6 RNAP (Fig. 3b and Supplementary Table 4). In unmodified RNAs synthesized with SP6 RNAPs, the rA → rG/dT → dC substitution was observed to be the predominant error (Fig. 3c and Supplementary Table 5). However, the substitution profiles of m1 ψ - or ψ -modified RNA demonstrated a preponderance of rA → rU/dT → dA substitutions as observed with m1 ψ - or ψ -modified RNAs synthesized with T7 RNAP (Fig. 3c and Supplementary Table 5). Compared to the error rates of unmodified RNA, ψ -modified RNAs had seven- to nine-fold increase in rA → rU/dT → dA substitution (Supplementary Table 5). For the m1 ψ -modified RNAs, the combined error rates were two- to three-fold higher than unmodified RNA. Furthermore, the sequence context analysis of the rA → rU substitutions observed in RNAs synthesized with SP6 RNAP demonstrated enrichment of a 5' rC for both unmodified and modified RNAs (Supplementary Fig. 5). Additionally, the substitution sites were distributed throughout the length of the transcript as observed in RNAs synthesized with T7 RNAP (Supplementary Fig. 6). Taken together, SP6 RNAP has a higher total error rate as compared to T7 RNAP but the error profile and the distribution of the error is comparable between these RNAPs.

Fidelity of uridine incorporation is not dependent on the yield of the in vitro transcription reaction. For synthetic mRNA-based applications, high yield of RNA from the in vitro transcription reaction is desirable and reactions are typically performed with high concentrations of rNTPs. The recommended high-yield rNTP concentrations are different for T7 RNAP (40 mM rNTP) and SP6 RNAP (20 mM rNTP). In order to ensure that the differences in combined error observed for T7 RNAP and SP6 RNAP are not due to differences in the rNTP concentrations in the reactions, we performed in vitro transcription reactions with T7 RNAP under low rNTP reaction conditions with either 20 mM or 10 mM rNTP (total). The total combined error rates as well as the base substitution errors observed in unmodified RNA1/RNA2 when reactions were performed with T7 RNAP under low rNTP (10 or 20 mM) reaction conditions were comparable to that observed with high rNTP (40 mM) reaction conditions (Supplementary Table 6) suggesting that the overall rNTP concentration in the reaction does not affect fidelity of uridine incorporation and the differences in error rates observed with T7 and SP6 RNAP are not due to differences in the reaction conditions.

Altering the rNTP composition during in vitro transcription reduces combined error rate and the predominant rA-to-rU substitution error. Our data demonstrates that m1 ψ - and ψ -modified RNAs have increased combined errors compared to unmodified RNAs and the increased rA \rightarrow rU substitution errors account for most of the misincorporations that are observed (Figs. 1c and 3c). Furthermore, the increased rA \rightarrow rU substitution errors observed in RNAs synthesized with SP6 RNAP suggest that the predominant rA \rightarrow rU substitution might occur during in vitro transcription where the uridine analogs are misincorporated with higher frequency. In order to test this hypothesis as well as to reduce the rA \rightarrow rU substitution during in vitro transcription reactions with uridine analogs, we decided to manipulate the rUTP concentration in the reaction with the idea that balancing the rUTP in the in vitro transcription reaction to match the nucleotide composition of the RNA sequence might result in reduced rA \rightarrow rU substitution errors and consequently the total errors observed during in vitro transcription.

One aspect of rationalized mRNA design that has been gaining traction has been to reduce the uridine composition without altering the amino acid sequence of the protein encoded from the synthetic mRNA^{29,30}. Uridine-depletion in Cas9 mRNA sequence demonstrated a reduction of innate immune response and an increase in Cas9 activity²⁹. Comirnaty and Spikevax sequences consist of 19% and 15% uridine, respectively, as compared to the wild-type spike protein sequence that has 33% uridine in the sequence^{30,31}. Often, rationalized design of the synthetic mRNA molecule is further combined with reaction optimization such as altering the rNTP concentrations in the reaction to optimize the RNA yields from the reactions as well as reduction of dsRNA byproducts^{12,18}.

First, as a control, we analyzed the error rates when in vitro transcription reactions were performed under standard high-yield rNTP condition where all the rNTPs are added equally to a final concentration of 40 mM (represented as equal in Fig. 4 and Supplementary Table 7). The uridine-depleted RNA sequences transcribed with T7 RNAP with equal molar unmodified rNTPs had a total combined error rate $6.0 \pm 1.1 \times 10^{-5}$ errors/base, while the error rates of ψ - and m1 ψ -incorporating transcripts were observed to be $1.9 \pm 0.4 \times 10^{-4}$ and $9.9 \pm 3.0 \times 10^{-5}$ errors/base, respectively (Fig. 4a and Supplementary Table 7). For the rUTP optimized reactions, no difference in the RNA yield from reactions with modified uridine analogs was observed (Supplementary Table 8). Furthermore, the integrity of the RNA was not compromised when reactions were performed with proportional rNTPs (Fig. 4b). The total combined error rate of unmodified RNAs was observed to be $5.6 \pm 2.1 \times 10^{-5}$ errors/base and that of m1 ψ -incorporating RNAs was $6.3 \pm 0.5 \times 10^{-5}$ errors/base (Supplementary Table 7). Noticeably, the total error rate of ψ -modified RNAs was $8.4 \pm 0.9 \times 10^{-5}$ errors/base, about one-fold reduced as compared to equal molar rNTP condition ($1.9 \pm 0.4 \times 10^{-4}$ errors/base).

Under standard equal molar rNTP reaction conditions, the substitution error profile of the uridine-depleted RNA sequences resembled those of RNA1/RNA2, CLuc mRNA and BNT162b2, with rA \rightarrow rU/dT \rightarrow dA substitution demonstrating the most significant change when modified uridine was used in the reaction (Figs. 1c, 4c,d; Supplementary Fig. 7; Supplementary Tables 9 and 10). For the RNA sequence with 5.5% uridine content, the rA \rightarrow rU/dT \rightarrow dA substitution was increased three-fold in m1 ψ -modified RNA as compared to the unmodified RNA and this was even more pronounced in the ψ -modified RNA with an increase of 14-fold over unmodified RNA. Interestingly, when the rNTP concentrations were altered to match the nucleotide content of the sequence, the rA \rightarrow rU substitution error rates were lowered significantly as compared to that observed under standard reaction condition (equal rNTP conditions). For unmodified RNA, a four-fold reduction in rA \rightarrow rU substitution was observed when rUTP concentration was altered to match the template sequence (Fig. 4c and Supplementary Table 9). Similarly, six- and seven-fold reductions were observed for ψ -modified and m1 ψ -modified RNA, respectively. Furthermore, no significant change in any of the other substitution errors were observed when the rNTP concentrations were altered (Fig. 4c and Supplementary Table 9). Similar fold changes were also observed for two other artificial RNA sequences with uridine content of 6.6% or 12.3% and U-depleted CLuc mRNA sequence (Fig. 4d; Supplementary Fig. 7; Supplementary Tables 9 and 10). The comparison of the two different RNA synthesis workflows with uridine-depleted RNA sequences demonstrates that lowering the rUTP concentration to match the nucleotide sequence of the RNA indeed reduced the rA \rightarrow rU substitution error during in vitro transcription reaction and further provides support that the rA \rightarrow rU substitutions observed with the uridine modifications are errors observed during in vitro transcription.

In light of our observation that the fidelity of T7 RNAP can be modulated by adjusting the rNTP ratio in the reaction, we investigated if the same holds true for SP6 RNAP where the rA \rightarrow rU substitution was also observed to be the most predominant substitution error in modified RNAs. We used the same T-depleted randomized sequence template (resulting in 5.5% uridine content in the RNA) and reactions were performed with either equimolar rNTPs or rNTPs matched to the nucleotide sequence. In presence of proportional rNTPs, the total combined error rate in unmodified RNA was reduced to $1.3 \pm 0 \times 10^{-4}$ errors/base, one-fold decrease from

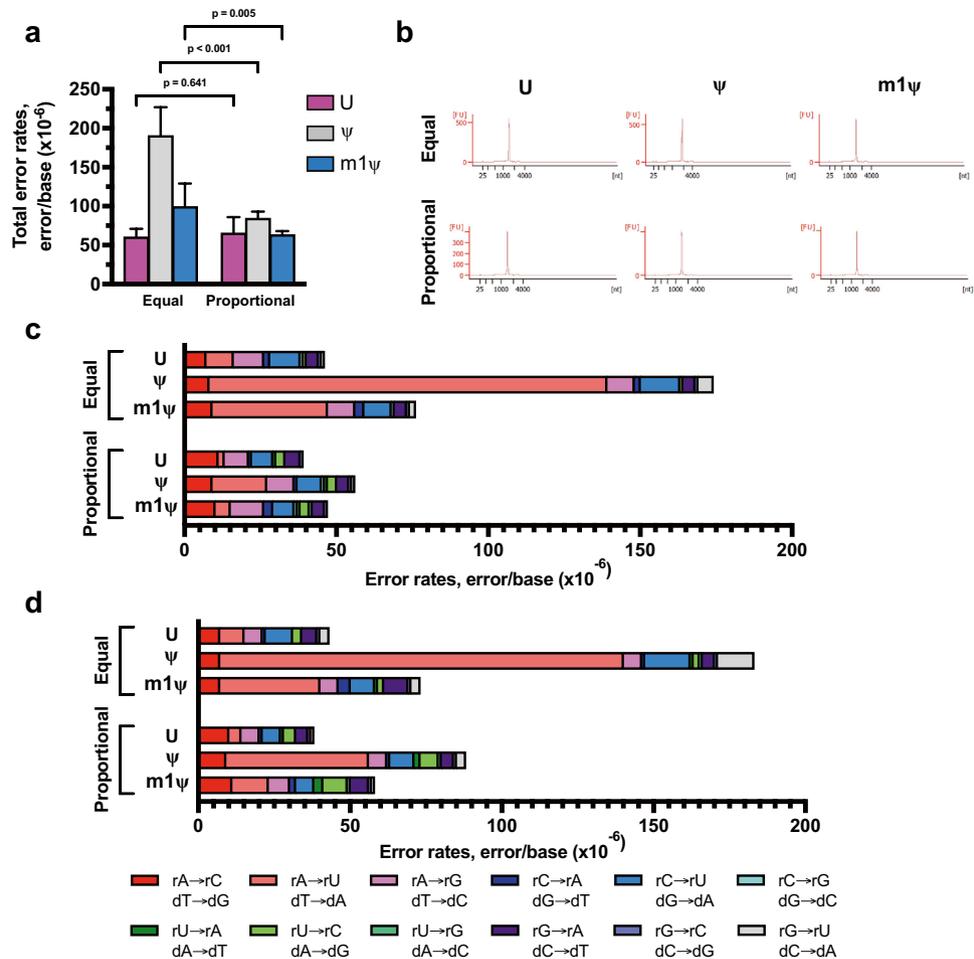


Figure 4. In vitro transcription error profile of T7 RNA polymerase can be modulated by altering ribonucleotide composition in the reaction. Uridine-depleted artificial RNA sequences and uridine-depleted CLuc mRNA were transcribed with T7 RNA polymerase and reactions were performed with either equal molar rNTPs or rNTPs proportional to the template sequence composition. **(a)** Combined first-strand errors are an average of four uridine depleted RNA sequences under the two different rNTP conditions. The three synthetic RNAs have uridine content of 5.5%, 6.6% and 12.3%; CLuc mRNA has a uridine content of 13.5%. Unpaired t-tests were used to calculate p-values. **(b)** Bioanalyzer traces of synthetic RNA with 5.5% uridine content demonstrating synthesis of full-length IVT products under both reaction conditions. **(c)** Base substitution error profile observed for unmodified and modified uridine-depleted artificial RNA sequences (RNA with 5.5% uridine content) under two different rNTP reaction conditions. **(d)** Base substitution error profile observed for unmodified and modified uridine-depleted CLuc mRNA from under the two different rNTP reaction conditions. Colors indicate specific substitution errors observed.

$2.8 \pm 0.9 \times 10^{-4}$ errors/base in presence of equal molar unmodified rNTP (Supplementary Table 11). Similarly, for m1 ψ -modified reaction, the total combined error rate was reduced two-fold (from $4.1 \pm 1.0 \times 10^{-4}$ errors/base to $1.5 \pm 0 \times 10^{-4}$ errors/base). As observed with T7 RNAP, ψ -modified RNA had the most pronounced (three-fold) reduction in total combined error rate (from $6.0 \pm 2.1 \times 10^{-4}$ errors/base to $1.7 \pm 0 \times 10^{-4}$ errors/base) when the rNTP ratios were altered to match the sequence of the RNA. Furthermore, the reduction in total combined error observed in proportional rNTP reaction conditions is attributable specifically to reduction in the rA \rightarrow rU substitution errors (Fig. 5 and Supplementary Table 12).

Uridine-depletion of the RNA sequence is not a viable alternative for all sequences. Furthermore, the extent of uridine-depletion is dependent on the sequence since it requires change in the uridine content without altering the codon. We investigated a corollary approach where we hypothesized that increasing the rATP concentrations in the reaction might also reduce the rA \rightarrow rU substitutions in the in vitro transcription reactions. Altering the rNTP ratio did not compromise the integrity of the full-length transcript (Fig. 6a). When 16 mM rATP was used in the reactions (with 8 mM each of the other three rNTPs), the total combined error rates of unmodified, ψ - and m1 ψ -modified RNA1/RNA2 were observed to be $4.5 \pm 0.5 \times 10^{-5}$ errors/base, $6.4 \pm 0 \times 10^{-5}$ errors/base and $6.0 \pm 0.9 \times 10^{-5}$ errors/base, respectively (Supplementary Table 13). When 20 mM rATP was used (with 10 mM each of the other three rNTPs), the total combined error rates of unmodified, ψ - and m1 ψ -modified RNA1/RNA2 were observed to be $4.7 \pm 0 \times 10^{-5}$ errors/base, $6.6 \pm 0.2 \times 10^{-5}$ errors/base and $4.8 \pm 0.7 \times 10^{-5}$ errors/base,

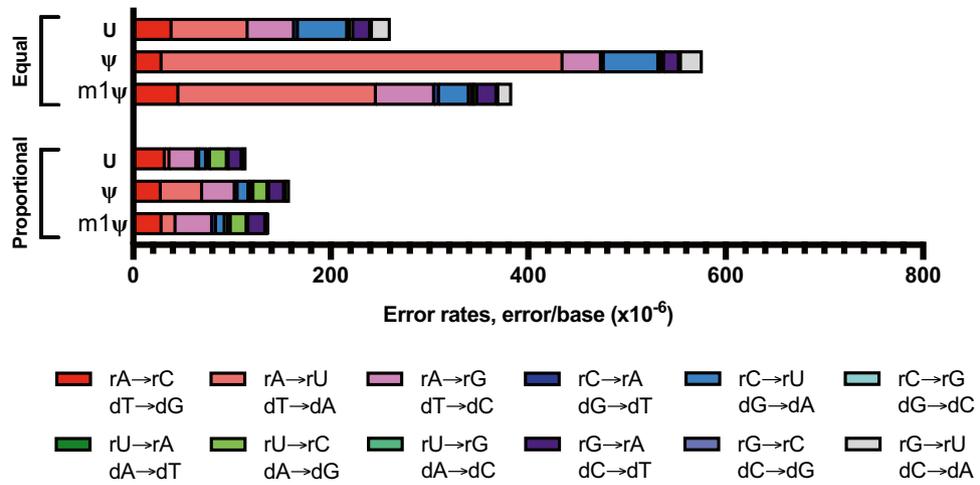


Figure 5. In vitro transcription error profile of SP6 RNA polymerase can be modulated by altering ribonucleotide composition in the reaction. Uridine-depleted artificial RNA sequence (uridine content of the sequence is 5.5%) was transcribed with SP6 RNA polymerase and reactions were performed with either equal molar rNTPs or rNTPs proportional to the template sequence composition. Base substitution error profile observed for unmodified and modified uridine-depleted RNA sequence under the two different rNTP reaction conditions. Colors indicate specific substitution errors observed.

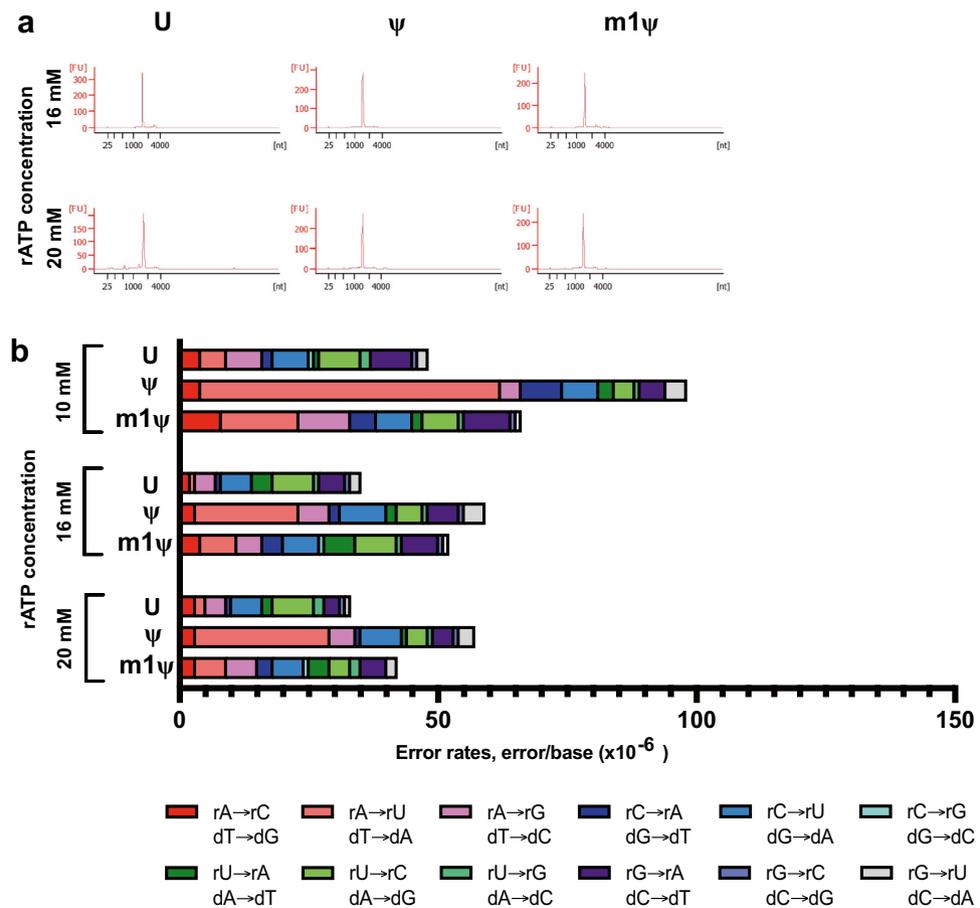


Figure 6. In vitro transcription error profile of T7 RNA polymerase can be modulated by increasing the rATP concentration in the reaction. Artificial RNA sequences RNA1 and RNA2 were transcribed with T7 RNA polymerase and reactions were performed with excess of rATP (16 mM or 20 mM) as compared to the other rNTPs in the reaction (8 mM or 10 mM respectively). (a) Bioanalyzer traces of RNA1 synthesized with either 16 mM or 20 mM rATP demonstrating synthesis of full-length IVT products. (b) Base substitution error profile observed for unmodified and modified RNA sequences under the different rATP concentrations.

respectively (Supplementary Table 13); a two-fold decrease in total combined error rates for ψ -modified RNAs. As hypothesized, the substitution error profile of reactions performed with excess rATP showed rA \rightarrow rU/dT \rightarrow dA substitution in the modified RNA changed the most compared to unmodified RNA (Fig. 6b and Supplementary Table 14).

Taken together, these results demonstrate that the rA \rightarrow rU substitution that is observed when in vitro transcription reactions are performed with uridine analogs, stems from substitution errors during in vitro transcription with T7 and SP6 RNAPs. Furthermore, the fidelity of the uridine analog incorporation can be increased by either lowering the rUTP concentration to match the nucleotide composition of the synthetic RNA sequence or increasing the rATP concentration in the reaction without compromising the yield from the reaction.

Discussion

The current approaches to introduce modifications in synthetic mRNAs involve replacing a canonical nucleotide with a modified analog during in vitro transcription by ssRNAPs. T7 and SP6 RNAPs are the most commonly used RNA polymerases for in vitro transcription. Both T7 and SP6 have their own promoter specificities and it is also well known that they result in heterogeneous RNA populations^{28,32,33}. However, whether or not these two RNA polymerases incorporate modified nucleotides with similar fidelities has not been tested till date. Our data demonstrates that the combined error rates in RNAs synthesized with SP6 RNAP were up to two-fold higher than those observed with T7 RNAP (Figs. 1 and 3; Table 1; Supplementary Tables 3 and 4). On the other hand, T3 RNAP exhibited error rates comparable to that of T7 (Supplementary Table 3). That said, it is interesting to note that the error profiles of uridine-modified RNA for both T7 and SP6 polymerases were similar. For both RNAPs, the incorporation of ψ or m1 ψ was more error prone than canonical uridine incorporation (Figs. 1a and 3a; Table 1; Supplementary Table 4). However, when uridine analogs were present in the reaction, both T7 and SP6 RNAPs had a very similar substitution error profile with rA \rightarrow rU substitution being the predominant error observed during in vitro transcription (Figs. 1c and 3c; Supplementary Tables 2 and 5). Available crystal structure capturing T7 RNAP during transcription elongation, as well as biochemical assays and molecular dynamics simulation, have identified several key residues that contribute to transcription fidelity^{34–36}. For example, Tyrosine 639 is known to be critical for the pre-insertion of correct nucleotides opposite to the template DNA strand³⁶. Since the differences in the total combined errors between T7 and SP6 RNAPs were observed for the same template sequences, and since base pairing between incoming rNTP and DNA template is the same, the fidelity differences of the two enzymes must be associated with differences in the protein sequence, e.g. residues which function to ensure correct base pairing. Interestingly, Tyr639 is conserved in both T7 and SP6 RNAPs as well as among other bacteriophage ssRNAPs, but neighboring amino acids are not. Future studies to elucidate the role of residues outside of the active site on fidelity of modified nucleotide incorporation will be instrumental in designing high-fidelity in vitro transcription systems. In addition to targeted amino acid replacements in T7 or SP6 RNAP, determining the error rates of homologous ssRNAPs will reveal the sequence determinants of transcriptional fidelity within this protein family. Furthermore, it will also be of interest to identify novel ssRNAPs that have a distinct substitution error profile than that of T7 RNAP and might be a better RNAP candidate for introduction of uridine analogs.

Even though the combined error rates observed for the same set of RNA sequences were significantly different between T7 and SP6 RNAPs, there were few similarities. First, the fidelity with which the analogs were incorporated, followed the same trend—uridine > m1 ψ > ψ —with m1 ψ -modified RNA exhibiting lower total errors and rA \rightarrow rU/dT \rightarrow dA substitution than ψ -modified RNA (Figs. 1 and 3; Supplementary Tables 2 and 5). Second, the difference in fidelity between uridine-containing RNAs and ψ /m1 ψ -modified RNAs is mainly attributable to a higher rA \rightarrow rU/dT \rightarrow dA substitution error (Figs. 1 and 3; Supplementary Tables 2 and 5). Based on our data from multiple templates and reaction conditions, we demonstrate that both ψ and m1 ψ have a higher propensity to mis-pair with dT in the template DNA during in vitro transcription. Biophysical studies to measure the melting temperatures (T_m) of synthetic RNA duplexes containing either uridine, ψ or m1 ψ , have shown that both ψ - or m1 ψ -containing duplexes have a higher T_m than uridine-containing duplexes¹³. Increased base pairing and stacking has been postulated to contribute to the increased T_m observed^{30,37–39}. One of the critical features of ψ and m1 ψ is the C5-C'1 bond that enables rotation between the sugar moiety and the nucleobase, which, in contrast to canonical uridine, could provide improved base pairing and stacking. As compared to uridine and m1 ψ , ψ contains an extra hydrogen bond donor group (N1H) that imparts a universal base character to ψ . Hence, ψ can not only pair with A but can also wobble base-pair with G, U, or C³⁶. On the other hand, m1 ψ has a methyl group in the N1-position and therefore does not have the extra hydrogen bond donor and therefore wobble pairing with other nucleotides is not favored. It is plausible that the lower error rates observed in m1 ψ -containing RNAs is due to the lack of the extra hydrogen bond donor³⁹. Additionally, it is likely that the methyl group in the N1 position of m1 ψ may alter the polarity of the C2 position and make it less favorable to pair with dT in the DNA template. It is also worth noting that stability difference between RNA duplex containing four consecutive ψ 's and m1 ψ 's was not distinguishable¹³. Stability measurements in which a single r ψ /TP or rm1 ψ /TP pairs with DNA template to mimic the errors observed during in vitro transcription will likely be more relevant to distinguish the base-pairing differences we observe in this study. Not much is understood about the pairing of r ψ or r m1 ψ with DNA, but there is precedence for RNA-RNA interactions in presence of ψ and m1 ψ which might help us understand the increased error rates observed in presence of the modified analogs^{13,38–40}.

Combining sequence optimization of the synthetic RNA with incorporation of uridine modifications, specifically m1 ψ and ψ , in synthetic mRNA-based vaccines and therapeutics is becoming a common practice. In addition, depleting the uridine content in the synthetic mRNA by sequence optimization has been demonstrated to reduce the immunogenicity of the synthetic molecules^{29,30}. Our data demonstrates that combining uridine depletion of the RNA sequence with altering the rNTP composition of the reaction, reduces the rA \rightarrow rU substitutions

that are introduced during in vitro transcription (Figs. 4c,d, 5; Supplementary Fig. 7; Supplementary Tables 9, 10, 12). Future studies aimed at comparing the outcome from these modified processes (uridine depletion with altered rNTP composition to generate high-fidelity synthetic mRNA with reduced dsRNA byproducts) will provide further insights into ways to optimize the efficacy from the synthetic mRNA drug substance.

Our observation that the prevalent rA → rU substitutions introduced during in vitro transcription can be reduced by balancing the rNTPs in the reaction suggests a few possible mechanisms for introduction of these substitution errors. By limiting the rψTP amounts or competing with excess rATP, the rA → rU substitution can be altered to reduce the error rates without affecting any of the other substitution errors (Fig. 6 and Supplementary Table 14). One of the possible reasons for this could be, that these mispairing events occur in presence of excess rNTPs in the reaction. This is further supported by the fact that in experiments where the template sequence was intentionally biased to have reduced uridine, adding equal molar rNTPs in the reaction (10 mM each) led to an overall increased error profile than that observed in templates that had comparable nucleotide usage (RNA1 and RNA2). Interestingly, the nucleotide composition of RNA1 and RNA2 have equal representation of all the four nucleotides and we still observed prevalent rA → rU substitutions when equal molar rNTPs were added in the reaction (Figs. 1c and 3c; Supplementary Tables 2 and 5). These reactions were performed with 10 mM rNTP each. It is possible that not all the rNTPs in the reaction are utilized by T7 RNAP there is always excess of rNTPs in these high-yield reaction conditions. Future studies to dissect out if the misincorporations are more prevalent in the early vs late stage of the in vitro transcription process will provide further insight into the events leading to uridine analog misincorporation. It can also be envisioned that in vitro transcription systems that limit the initial rNTP concentration but allow for a steady, optimized rNTP feeding mechanism might further help improve the fidelity of nucleotide incorporation and be instrumental in generating high-fidelity synthetic mRNAs.

In order to achieve high-fidelity RNA products, it is desirable to understand the rules of nucleotide incorporation so that if there are sequences that are more error prone, they can be omitted from the synthetic mRNA during design. However, in this study, the comparison of multiple sequence contexts demonstrated that other than T7 RNAP-incorporated ψ-modified RNAs that demonstrated slight sequence context preference, there is no strong correlation between the sequence context of the DNA template and the substitution errors observed under any condition tested (Fig. 2a; Supplementary Figs. 4a and 5). Furthermore, the errors observed were distributed throughout of the length of RNA (Fig. 2b; Supplementary Figs. 4b and 6).

Our results demonstrate that the presence of ψ and m1ψ in the in vitro transcription reactions result in higher base substitution errors in the modified RNAs. A critical aspect to consider is how these errors might affect the efficacy of the synthetic mRNA drug substance and how much of these error-prone molecules can be tolerated in vivo without any adverse effect. Because in vitro transcribed mRNAs are modified throughout the body of the mRNA, it is also critical to consider if these mRNAs are faithfully translated in the cell. We have previously demonstrated that in human embryonic kidney cell, low frequency translation elongation miscoding events are observed from ψ-containing mRNAs due to altered tRNA selection in ψ-containing codons²². For m1ψ-substituted RNAs, it has been shown that translation initiation and ribosome transit is altered in vivo¹⁴. A complete understanding of what errors are incorporated during the RNA synthesis process and how that further affects the identity of the protein synthesized is a timely question given that therapeutic applications would require repeat dosing of the mRNAs and would also require expression of the protein of choice. To be able to predict the best outcome from these synthetic molecules, it is critical that we understand where variability comes from and to be able to define the rules to avoid these variabilities.

Materials and methods

Oligonucleotides and DNA template sequences used in this study. All of the oligonucleotides for in vitro transcription and reverse transcription were synthesized by Integrated DNA Technologies (IDT, Coralville IA) and the sequences are available in Supplementary Table 16. The DNA template sequences for in vitro transcription are available in Supplementary Table 17.

Generation of DNA templates for in vitro transcription (IVT) of long RNAs. Generation of the DNA templates for the artificial RNA sequences RNA1 and RNA2 (1122- and 1124-nucleotide sequence that includes all possible four-base combinations) were described earlier²⁶. For in vitro transcription reactions with SP6 or T3 RNA polymerase, the corresponding promoter sequences were inserted in the DNA templates using Q5 Site-Directed Mutagenesis Kit (E0554, New England Biolabs). DNA templates encoding functional mRNAs, *Cypridina* luciferase (CLuc) mRNA (1707 nucleotides) and part of BNT162b2/Comirnaty mRNA (4187 nucleotides)³⁰, were synthesized by GenScript Inc. (GenScript, Piscataway NJ) and introduced into standard high-copy plasmids. DNA templates for uridine-depleted artificial sequences (with 5.5%-, 6.6%- and 12.3%-T content) were synthesized by GenScript and introduced into standard high-copy plasmids. The plasmids were propagated in *E. coli* (C2987, New England Biolabs) and purified with the Monarch Plasmid Miniprep Kit (T1010, New England Biolabs). T-depleted (13.5% T-content) version of the *Cypridina* luciferase (CLuc) mRNA template was synthesized by GenScript and introduced into standard high-copy plasmids. The plasmids were propagated in *E. coli* (C2987, New England Biolabs) and purified with the Monarch Plasmid Miniprep Kit (T1010, New England Biolabs). Plasmids were digested with restriction enzymes to generate linearized templates for in vitro transcription. The linearized plasmids were treated with PreCR Repair Mix (M0309, New England Biolabs) and purified with the Monarch PCR & DNA Cleanup Kit (T1030, New England Biolabs).

For short RNAs (30-nucleotide and 60-nucleotide long), double-stranded DNA templates were generated by annealing DNA oligonucleotides for template and non-template strands in annealing buffer (10 mM Tris pH 7.5, 50 mM NaCl, 1 mM EDTA) by heating at 95 °C for 5 min followed by gradual cooling to room temperature.

DNA templates corresponding to three uridine-depleted artificial RNA sequences with base composition of (1) 30.9% A, 33.4% C, 30.1% G, 5.5% U, (2) 29.7% A, 32.0% C, 31.7% G, 6.6% U and (3) 30.2% A, 30.4% C, 27.1% G, 12.3% U were synthesized by GenScript Inc. (GenScript, Piscataway NJ). Additionally, to evaluate if this approach is viable for a functional mRNA, we also generated a uridine-depleted CLuc mRNA sequence where the uridine content was reduced from 22.8 to 13.5%. The uridine-depletion of the CLuc mRNA sequence was performed without perturbing the amino acid sequence of the encoded protein so that there are no other differences between the two mRNAs, other than the nucleotide content of the sequences (Supplementary Table 17).

In vitro transcription (IVT). In vitro transcription reactions were performed according to standard protocols provided with the high-yield in vitro transcription kits (E2040 and E2070, New England Biolabs), consisting of 40 mM rNTP (pH buffered with sodium phosphate) for T7 RNA polymerase and 20 mM rNTP (pH buffered with Tris) for SP6 RNA polymerase. In vitro transcription with T3 RNA polymerase (M0378S, New England Biolabs) was carried out following manufacturer's instruction. For modified RNAs, UTP was replaced with either pseudouridine-5'-triphosphate (N-1019, TriLink Biotechnologies) or N¹-Methylpseudouridine-5'-Triphosphate (N-1081, TriLink Biotechnologies). For the uridine-depleted sequences (RNA with 5.5%, 6.6% and 12.3% uridine content), two sets of reactions were performed—one in which all the rNTPs were present in equimolar ratio and another in which the rNTPs were altered to match the template sequence composition. Following IVT, the DNA template was digested with Turbo DNase (AM2238, Invitrogen) digestion at 37 °C for 30 min and then purified with the Monarch RNA Cleanup Kit (T2050, New England Biolabs) for long RNA (1020 nucleotides to 4187 nucleotides) or with the Oligo Clean-Up and Concentration Kit (34100, Norgen Biotek Inc) for short 30-nt RNA and 60-nt RNA.

Low-yield IVT reaction conditions were carried out with either 20 mM or 10 mM rNTPs (total) at 37 °C for two hours. DNA template removal and cleanup were performed as described above.

For the rUTP optimized T7 RNAP reactions, rNTPs proportional to the template sequence were used [for e.g., 12.4 mM (31%) rATP, 13.2 mM (33%) rCTP, 12 mM (30%) rGTP and 2.4 mM (6%) rUTP was used for RNA sequence with 5.5% uridine content]. For the rUTP optimized SP6 RNAP reactions, rNTPs proportional to the template sequence were used (for e.g., 6.2 mM (31%) rATP, 6.6 mM (33%) rCTP, 6 mM (30%) rGTP and 1.2 mM (6%) rUTP was used for RNA sequence with 5.5% uridine content].

In vitro transcription of RNA1/RNA2 with excess rATPs were performed with either 16 mM rATP with 8 mM of other rNTPs or 20 mM rATP with 10 mM of other rNTPs.

Bioanalyzer for RNA size distribution and integrity. Eluted RNA samples from the in vitro transcription reactions were diluted based on concentrations measured on a Nanodrop spectrophotometer (13-400-519, Thermo Fisher Scientific) and denatured at 70 °C for 2 min and snap-cooled on ice. 250 ng RNA samples were analyzed with RNA 6000 Nano kits (5067, Agilent Technologies) and the integrity and size distribution of the RNA was assessed using mRNA Nano series 2 assay (G2938, Agilent Technologies).

Nucleoside digestion of RNA and UHPLC-MS analyses to assess modification incorporation. Purified modified RNA and RNA without any chemical modification were digested with the Nucleoside digestion mix (M0649, New England Biolabs) at 37 °C for 1 h. Base composition analysis was performed by Liquid Chromatography-Mass Spectrometry (LC-MS) using an Agilent 1290 Infinity II UHPLC equipped with G7117A Diode Array Detector and 6135XT MS Detector, on a Waters Xselect HSS T3 XP column (2.1 × 100 mm, 2.5 μm) with the gradient mobile phase consisting of methanol and 10 mM ammonium acetate buffer (pH 4.5).

Intact mass spectroscopy analysis for short RNA oligonucleotides. Purified 30-nucleotide and 60-nucleotide RNA with different uridine modifications were analyzed at Novatia, LLC using on-line desalting, flow injection electrospray ionization on an LTQ-XL ion trap mass spectrometer (Thermo Fisher Scientific) and analyzed with ProMass Deconvolution software from Novatia, LLC (https://www.enovatia.com/our-products/promass/?gclid=EAIaIQobChMIoZS6xKDn-AIViNrICh3peATrEAAYASAAEgLRNPD_BwE).

First and second strand cDNA synthesis for PacBio. The cDNA synthesis was performed as described before with a modified cleanup step using Monarch PCR & DNA Cleanup Kit (T1030, New England Biolabs)²⁶. Sequences of forward and reverse oligonucleotides used are provided in the Table S1.

Pacific biosciences SMRTbell library preparation and sequencing. The library preparation for sequencing on RSII system was performed as described before²⁶. For the sequencing on the Sequel platform, about 1.5 μg cDNA was treated with NEBNext End Repair Module (E6050, New England Biolabs) at room temperature for 5 min, followed by purification with Monarch PCR & DNA Cleanup Kit (T1030, New England Biolabs). The end-repaired cDNA was ligated with 2 μL barcoded adaptor (100-466-000, Pacific Biosciences) with T4 DNA Ligase (M0202, New England Biolabs) in 50 μL reaction volume at room temperature for 1 h, followed by purification with Monarch PCR & DNA Cleanup Kit (T1030, New England Biolabs). The un-ligated adaptor and cDNA were digested with *E. coli* Exonuclease III (M0206, New England Biolabs) and Exonuclease VII (M0379, New England Biolabs) in 1X standard Taq buffer at 37 °C for 1 h, followed by cleaning up with Monarch PCR & DNA Cleanup Kit (T1030, New England Biolabs). The ligated DNA was repaired with PreCR Repair Mix (M0309, New England Biolabs) at 37 °C for 30 min. The libraries were purified with 0.6X volume of AMPure PB beads (100-265-900, Pacific Biosciences) and pooled for sequencing runs. SMRT Link was used to generate the

protocol for primer annealing, polymerase binding [Sequel Binding Kit 3.0 (101-613-900, Pacific Biosciences)], cleanup and final loading to SMRT Cells LR and sequencing using Sequel system.

Data analysis. Analysis of sequencing data was performed as previously described²⁶. In short, high-accuracy consensus sequences were built for the first and second strand for each sequenced double-stranded DNA. The consensus sequences were aligned to the reference sequence and base substitutions, deletions, and insertions were determined. The first strand error rates were determined by comparing the first strand consensus sequences to the reference sequence (RNA strand), and mutations were required to be present in both strands. The average error and standard deviation of total errors were then calculated by combining the data from the templates for each reaction condition: for RNA1 and RNA2, one measurement was performed and for other RNA sequences, two independent repeats were performed. Substitution error profiles are represented separately for each RNA sequence. The relative fold change was calculated for each substitution as $(M - U)/U$, where M is the substitution rate on modified RNA ($m1\psi$ or ψ) and U is the substitution rate on unmodified RNA. For comparing total combined error rates and for estimating the statistical significances, p-values were calculated using unpaired t-tests from the measurements of unmodified and modified RNAs, and those under equal and proportional rNTP conditions. The local context of base substitutions was extracted, and sequence logos were built using WebLogo software⁴¹. Also, the frequency of mutations at each position was plotted to examine distribution of errors along the reference sequence.

Data availability

Sequencing data has been deposited into the Sequencing Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) under accession number SRP375186.

Received: 6 June 2022; Accepted: 22 July 2022

Published online: 29 July 2022

References

- Sahin, U., Kariko, K. & Tureci, O. mRNA-based therapeutics—Developing a new class of drugs. *Nat. Rev. Drug Discov.* **13**, 759–780. <https://doi.org/10.1038/nrd4278> (2014).
- Pardi, N., Hogan, M. J., Porter, F. W. & Weissman, D. mRNA vaccines—A new era in vaccinology. *Nat. Rev. Drug Discov.* **17**, 261–279. <https://doi.org/10.1038/nrd.2017.243> (2018).
- Chaudhary, N., Weissman, D. & Whitehead, K. A. mRNA vaccines for infectious diseases: Principles, delivery and clinical translation. *Nat. Rev. Drug Discov.* **20**, 817–838. <https://doi.org/10.1038/s41573-021-00283-5> (2021).
- Baden, L. R. *et al.* Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N. Engl. J. Med.* **384**, 403–416. <https://doi.org/10.1056/NEJMoa2035389> (2021).
- Polack, F. P. *et al.* Safety and efficacy of the BNT162b2 mRNA COVID-19 vaccine. *N. Engl. J. Med.* **383**, 2603–2615. <https://doi.org/10.1056/NEJMoa2034577> (2020).
- Anderson, B. R. *et al.* Nucleoside modifications in RNA limit activation of 2'-5'-oligoadenylate synthetase and increase resistance to cleavage by RNase L. *Nucleic Acids Res.* **39**, 9329–9338. <https://doi.org/10.1093/nar/gkr586> (2011).
- Anderson, B. R. *et al.* Incorporation of pseudouridine into mRNA enhances translation by diminishing PKR activation. *Nucleic Acids Res.* **38**, 5884–5892. <https://doi.org/10.1093/nar/gkq347> (2010).
- Andries, O. *et al.* N(1)-methylpseudouridine-incorporated mRNA outperforms pseudouridine-incorporated mRNA by providing enhanced protein expression and reduced immunogenicity in mammalian cell lines and mice. *J. Control Release* **217**, 337–344. <https://doi.org/10.1016/j.jconrel.2015.08.051> (2015).
- Baiersdorfer, M. *et al.* A facile method for the removal of dsRNA contaminant from in vitro-transcribed mRNA. *Mol. Ther. Nucleic Acids* **15**, 26–35. <https://doi.org/10.1016/j.omtn.2019.02.018> (2019).
- Kariko, K., Buckstein, M., Ni, H. & Weissman, D. Suppression of RNA recognition by Toll-like receptors: The impact of nucleoside modification and the evolutionary origin of RNA. *Immunity* **23**, 165–175. <https://doi.org/10.1016/j.immuni.2005.06.008> (2005).
- Kariko, K. *et al.* Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. *Mol. Ther.* **16**, 1833–1840. <https://doi.org/10.1038/mt.2008.200> (2008).
- Nelson, J. *et al.* Impact of mRNA chemistry and manufacturing process on innate immune activation. *Sci. Adv.* **6**, eaaz6893. <https://doi.org/10.1126/sciadv.aaz6893> (2020).
- Parr, C. J. C. *et al.* N1-Methylpseudouridine substitution enhances the performance of synthetic mRNA switches in cells. *Nucleic Acids Res.* **48**, e35. <https://doi.org/10.1093/nar/gkaa070> (2020).
- Svitkin, Y. V. *et al.* N1-methyl-pseudouridine in mRNA enhances translation through eIF2alpha-dependent and independent mechanisms by increasing ribosome density. *Nucleic Acids Res.* **45**, 6023–6036. <https://doi.org/10.1093/nar/gkx135> (2017).
- Motorin, Y. & Helm, M. RNA nucleotide methylation. *Wiley Interdiscip. Rev. RNA* **2**, 611–631. <https://doi.org/10.1002/wrna.79> (2011).
- Wurm, J. P. *et al.* Identification of the enzyme responsible for N1-methylation of pseudouridine 54 in archaeal tRNAs. *RNA* **18**, 412–420. <https://doi.org/10.1261/rna.028498.111> (2012).
- Wurm, J. P. *et al.* The ribosome assembly factor Nep1 responsible for Bowen-Conradi syndrome is a pseudouridine-N1-specific methyltransferase. *Nucleic Acids Res.* **38**, 2387–2398. <https://doi.org/10.1093/nar/gkp1189> (2010).
- Hadas, Y. *et al.* Optimizing modified mRNA in vitro synthesis protocol for heart gene therapy. *Mol. Ther. Methods Clin. Dev.* **14**, 300–305. <https://doi.org/10.1016/j.omtm.2019.07.006> (2019).
- Corbett, K. S. *et al.* SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. *Nature* **586**, 567–571. <https://doi.org/10.1038/s41586-020-2622-0> (2020).
- Sample, P. J. *et al.* Human 5' UTR design and variant effect prediction from a massively parallel translation assay. *Nat. Biotechnol.* **37**, 803–809. <https://doi.org/10.1038/s41587-019-0164-5> (2019).
- Mauger, D. M. *et al.* mRNA structure regulates protein expression through changes in functional half-life. *Proc. Natl. Acad. Sci. USA* **116**, 24075–24083. <https://doi.org/10.1073/pnas.1908052116> (2019).
- Eyler, D. E. *et al.* Pseudouridylation of mRNA coding sequences alters translation. *Proc. Natl. Acad. Sci. USA* **116**, 23068–23074. <https://doi.org/10.1073/pnas.1821754116> (2019).
- Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA modifications in gene expression regulation. *Cell* **169**, 1187–1200. <https://doi.org/10.1016/j.cell.2017.05.045> (2017).

24. Roy, B. Effects of mRNA modifications on translation: An overview. *Methods Mol. Biol.* **2298**, 327–356. https://doi.org/10.1007/978-1-0716-1374-0_20 (2021).
25. Pardi, N., Muramatsu, H., Weissman, D. & Kariko, K. In vitro transcription of long RNA containing modified nucleosides. *Methods Mol. Biol.* **969**, 29–42. https://doi.org/10.1007/978-1-62703-260-5_2 (2013).
26. Potapov, V. *et al.* Base modifications affecting RNA polymerase and reverse transcriptase fidelity. *Nucleic Acids Res.* **46**, 5753–5763. <https://doi.org/10.1093/nar/gky341> (2018).
27. Potapov, V. & Ong, J. L. Examining sources of error in PCR by single-molecule sequencing. *PLoS ONE* **12**, e0169774. <https://doi.org/10.1371/journal.pone.0169774> (2017).
28. Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**, 8783–8798. <https://doi.org/10.1093/nar/15.21.8783> (1987).
29. Vaidyanathan, S. *et al.* Uridine depletion and chemical modification increase Cas9 mRNA activity and reduce immunogenicity without HPLC purification. *Mol. Ther. Nucleic Acids* **12**, 530–542. <https://doi.org/10.1016/j.omtn.2018.06.010> (2018).
30. Nance, K. D. & Meier, J. L. Modifications in an emergency: The role of N1-methylpseudouridine in COVID-19 vaccines. *ACS Cent. Sci.* **7**, 748–756. <https://doi.org/10.1021/acscentsci.1c00197> (2021).
31. Jeong, D. M., Artiles, K., Ilbay, O., Fire, A., Nadeau, K., Park, H., Betts, B., Boyd, S., Hoh, R., Shoura, M. Assemblies of putative SARS-CoV2 spike encoding mRNA sequences for vaccines BNT-162b2 and mRNA-1273. <https://github.com/NAalytics/Assemblies-of-putative-SARS-CoV2-spikeencoding-mRNA-sequences-for-vaccines-BNT-162b2-and-mRNA-1273> (2021).
32. Melton, D. A. *et al.* Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12**, 7035–7056. <https://doi.org/10.1093/nar/12.18.7035> (1984).
33. Stump, W. T. & Hall, K. B. SP6 RNA polymerase efficiently synthesizes RNA from short double-stranded DNA templates. *Nucleic Acids Res.* **21**, 5480–5484. <https://doi.org/10.1093/nar/21.23.5480> (1993).
34. Huang, J., Briebe, L. G. & Sousa, R. Misincorporation by wild-type and mutant T7 RNA polymerases: Identification of interactions that reduce misincorporation rates by stabilizing the catalytically incompetent open conformation. *Biochemistry* **39**, 11571–11580. <https://doi.org/10.1021/bi000579d> (2000).
35. Temiakov, D. *et al.* Structural basis for substrate selection by t7 RNA polymerase. *Cell* **116**, 381–391. [https://doi.org/10.1016/S0092-8674\(04\)00059-5](https://doi.org/10.1016/S0092-8674(04)00059-5) (2004).
36. Duan, B., Wu, S., Da, L. T. & Yu, J. A critical residue selectively recruits nucleotides for t7 RNA polymerase transcription fidelity control. *Biophys. J.* **107**, 2130–2140. <https://doi.org/10.1016/j.bpj.2014.09.038> (2014).
37. Davis, D. R. Stabilization of RNA stacking by pseudouridine. *Nucleic Acids Res.* **23**, 5020–5026. <https://doi.org/10.1093/nar/23.24.5020> (1995).
38. Westhof, E. Pseudouridines or how to draw on weak energy differences. *Biochem. Biophys. Res. Commun.* **520**, 702–704. <https://doi.org/10.1016/j.bbrc.2019.10.009> (2019).
39. Morais, P., Adachi, H. & Yu, Y. T. The critical contribution of pseudouridine to mRNA COVID-19 vaccines. *Front. Cell Dev. Biol.* **9**, 789427. <https://doi.org/10.3389/fcell.2021.789427> (2021).
40. Kierzek, E. *et al.* The contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Res.* **42**, 3492–3501. <https://doi.org/10.1093/nar/gkt1330> (2014).
41. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190. <https://doi.org/10.1101/gr.849004> (2004).

Acknowledgements

The authors would like to thank Dillon Nye, Guy Birkenmeier, George Tzertzinis for comments on the manuscript and Rick Morgan, Sean Maguire, Alexey Fomenkov for help with the PacBio sequencing.

Author contributions

B.R. and T.H.C. conceived and designed the experiments; T.H.C. and N.D. carried out the experiments; V.P. and T.H.C. analyzed the data; J.L.O. and B.R. contributed reagents; B.R. and T.H.C. wrote the manuscript with input from all authors. All authors reviewed the manuscript.

Funding

This work was funded by New England Biolabs, Inc. Funding for open access charge: New England Biolabs, Inc.

Competing interests

Part of this work has been filed in a patent application (UPSTO provisional application number is 63/328654). T.H.C., V.P., N.D., J.L.O., and B.R. are employees of New England Biolabs Inc.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-17249-1>.

Correspondence and requests for materials should be addressed to B.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022